Evaluation of Latent Fingerprint Technologies (ELFT)

Phase II Test Results

Michael Indovina
mindovina@nist.gov
(301) 975-2927

Sponsors

- Department of Homeland Security (DHS S&T)
- Federal Bureau of Investigation (FBI CJIS)

Terminology

- **AFEM** = Automated Feature Extraction & Matching
- **SDK** = Software library encapsulating AFEM functionality
- **Hit rate = Identification rate =** Detection rate
- **Mate** = Exemplar = True mate = True match
- **Candidate** = an *alleged* mate
- **Non-mate** =incorrect candidate = False match/positive
- **Candidate list** = a list of candidates ordered by confidence
- **Rank** = relative position on the candidate list (rank 1 = “top”)
- **Hit** = mate on the candidate list (> threshold, e.g. rank)
- **Miss** = no mate on candidate list, or mate is < threshold

Approaches to Evaluating Latent Fingerprint Technology

- AFEM
- Extended Feature Sets (EFS)

Activities to Date

- **ELFT Phase I (2007)**
  - AFEM proof-of-concept (small dataset)
  - 11 participants (anonymous)
  - Aggregate (public) + individual (private) reports
- **ELFT Phase II (2008-present)**
  - AFEM with larger, operational data
  - 8 Participants from Phase I, submitted new SDKs
  - Final Report (in editorial review)
- **ELFT-EFS (*new*)**
  - More on this later…

ELFT Phase II Objectives

Measure & Characterize:

- **Accuracy of AFEM based systems**
  - Which factors contribute to errors they make ? (FP&FN)
- **Workload reduction**
  - Is the mate near the “top” ? (CMC, “cost metrics”)
  - Can we automatically eliminate false-matches? (e.g. thresholding)

ELFT Phase II Overview

- Tested 8 SDK's (one per participant), using
- Operational images from successful ***feature*** searches (IAFIS)
- Executed ***image-only*** searches to measure general AFEM accuracy
- Evaluated efficacy of candidate list reduction

Phase II Dataset: Databases

1) 100,000 rolled fingerprints (no mates –all *background*)
2) 100,000 rolled fingerprints ("seeded" mates + background)
3) 50,000 rolled fingerprints (no mates –all *background*)
4) 50,000 rolled fingerprints ("seeded" mates + background)

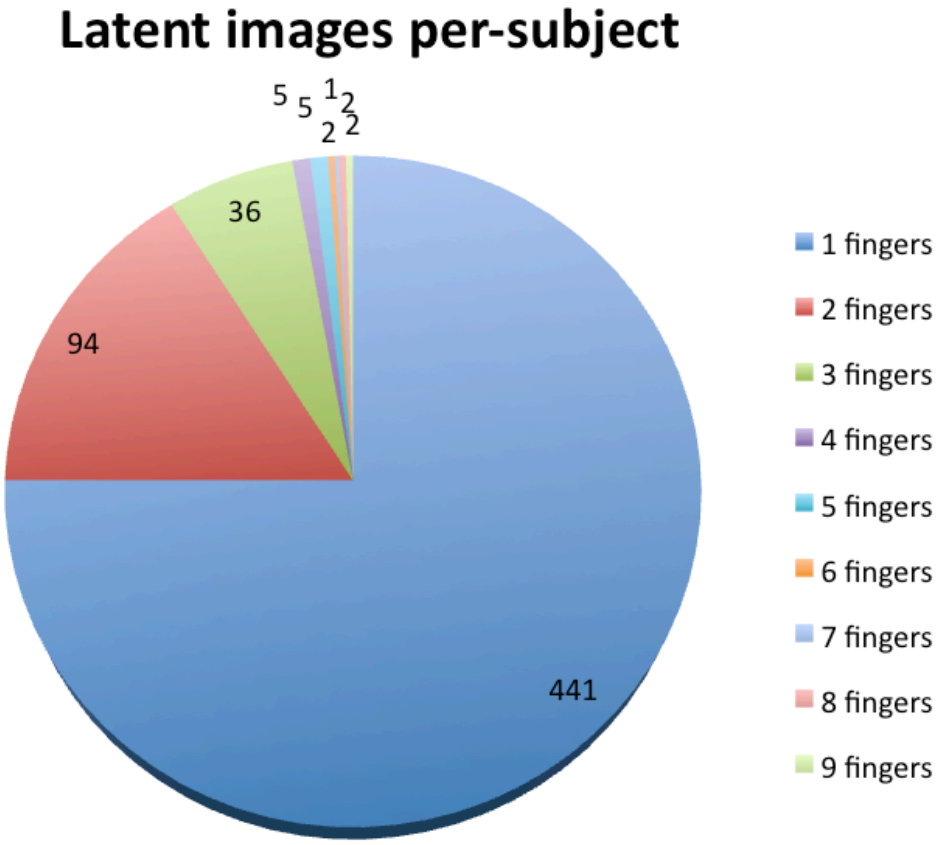
- *Background* from 4 operational sources
- 500 ppi
- 50% inked 50% live-scanned
- WSQ compressed 15:1

Phase II Dataset: Latents

1. **835 1000 ppi (native) images**
2. *835 500 ppi (sub-sampled) images*
3. *835 Region-of-Interest (ROI) "bit-mask" images*

- Casework over 2 to 3 year period
- Impression type = "Latent Photo"
- Paper source
- Correspond to 588 unique subjects

Phase II Dataset: Latents



Phase II Dataset: Latents

	ELFT Phase II Dataset (%)	FBI CMF (%)
Loops	46.8	65
Whorls	41.7	30
Arches	3.6	5
Undetermined	7.9	N/A

Finger position	Right thumb	Right index	Right middle	Right ring	Right little	Left thumb	Left index	Left middle	Left ring	Left little
% of total	29.6	15.2	7.7	2.4	0.7	17.6	12.1	7.7	5.7	1.3

Phase II Testing Protocol

1. 1000 ppi latents vs. 100,000 rolled database x 2
2. 1000 ppi latents vs. 50,000 rolled database x 2
3. 1000 ppi latents + ROI vs. 50,000 rolled databases x 2
4. 500 ppi latents vs 50,000 rolled databases x 2

- Directly compares effect of image resolution
- Directly compares database size (scalability)
- Directly compares effect of Region-of-Interest markup

Phase II Analysis Results

- **Overall Accuracy**
- **Effect of Database Size**
- **Effect of Resolution**
- **Effect of ROI**
- **Effect of Minutiae Count**
- Effect of Finger Position
- Effect of Pattern Class
- Execution Times
- ***Candidate List Fusion (multi-image, multi-algorithm)***

Metrics

- **Rank-based**
  - Number of searches resulting in the true-match ("mate") appearing on the candidate list.
  - Position ("ranks") at which they appear.
- **Score-based**
  - Number of searches resulting in false-positives ("non-mates") appearing on the candidate list above a specified "score threshold."

Rank-based Metrics

- **Method**
  - Each latent searched has one (and only one) "mate" (closed set)
- **Benefit**
  - Measures "identification rate" at various candidate list sizes
- **Disadvantage**
  - Doesn't consider frequency of false-positives when no mate exists in the database
- **Metric of choice**
  - CMC (Cumulative Match Characteristic) curves

Score-based Metrics

- **Method**
  - False-positives measured for searches without "mates" (open set). False-negatives ("misses") measured for searches with "mates" (at various FP thresholds).
- **Benefit**
  - Estimates Workload vs. Accuracy trade-off (Type I vs. II)
  - Thresholding useful for: candidate list reduction, "lights out" detections (watch list), and ULF
- **Metric of choice**
  - DET (Detection Error Trade-off) curves

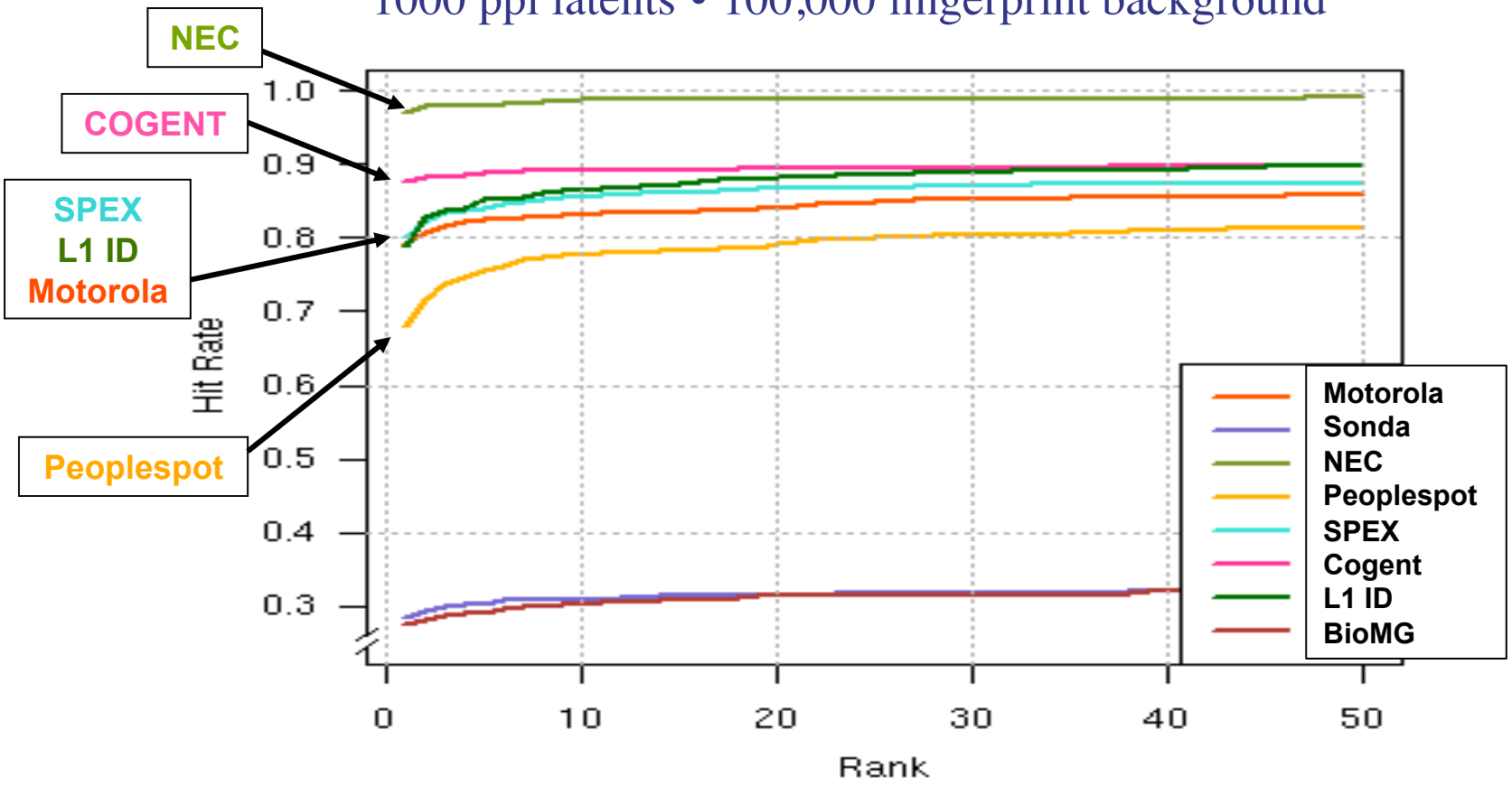
Main Results: CMC Curves

A Cumulative Match Characteristic (CMC) curve plots the probability of identification against the returned 1:N candidate list size.

It shows the probability that a given user appears in different sized candidate lists. The faster the CMC curve approaches 1, indicating that the user always appears in the candidate list of specified size, the better the matching algorithm.

Cumulative Match Characteristic (CMC)

1000 ppi latents • 100,000 fingerprint background



Detection Rates (**Rank 1**)

1000 ppi latents • 100,000 fingerprint background

Technology Provider	Identification Rate at Rank 1
NEC	97.2
Cogent	87.8
SPEX	80.0
Motorola	79.3
L1 Identity Solutions	78.8
Peoplespot	67.9
Sonda	28.5
BioMG	27.5

μ = 82%

Detection Rates (**Rank 10**)

1000 ppi latents • 100,000 fingerprint background

Technology Provider	Identification Rate at Rank 10	
NEC	98.8	+1.6
Cogent	89.2	+1.4
L1 Identity Solutions	86.5	+7.7
SPEX	85.6	+5.6
Motorola	83.2	+3.9
Peoplespot	77.8	+9.9
Sonda	30.9	
BioMG	30.2	**μ = 87%**

Detection Rates (**Rank 20**)

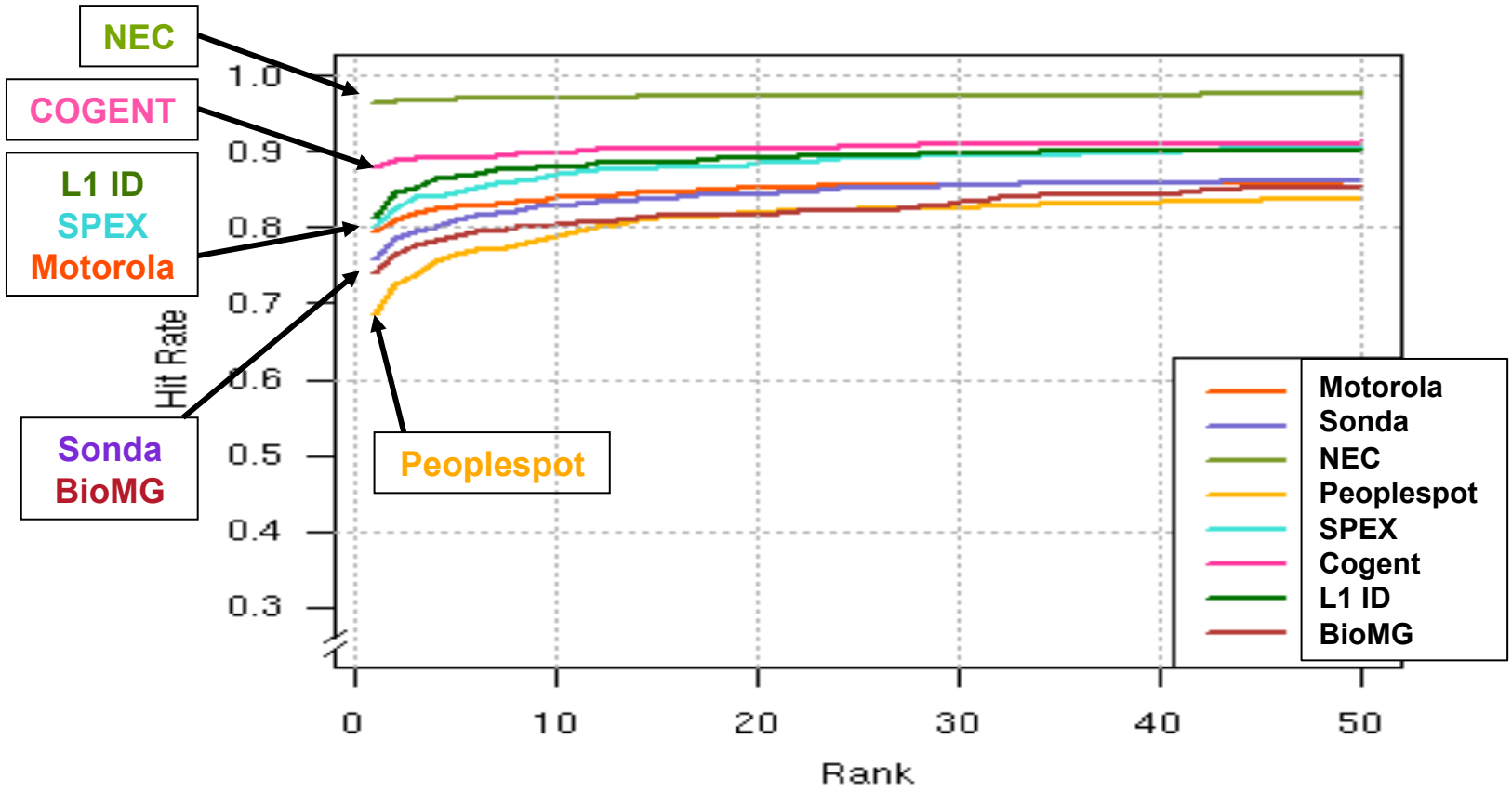
1000 ppi latents • 100,000 fingerprint background

Technology Provider	Identification Rate at Rank 20	
NEC	98.8	
Cogent	89.5	+0.3
L1 Identity Solutions	88.3	+1.8
SPEX	86.8	+1.2
Motorola	84.2	+1.0
Peoplespot	79.2	+1.4
Sonda	31.4	
BioMG	31.2	

μ = 88%

Cumulative Match Characteristic (CMC)

500 ppi latents • 50,000 fingerprint background



DET Curves

A DET curve plots error rates on both axes, giving uniform treatment to both types of error. The graph can then be plotted using logarithmic axes. This spreads out the plot and distinguishes different well-performing systems more clearly.

A DET curve plots the Type I (FPIR) vs. Type II (FNIR) error rate.

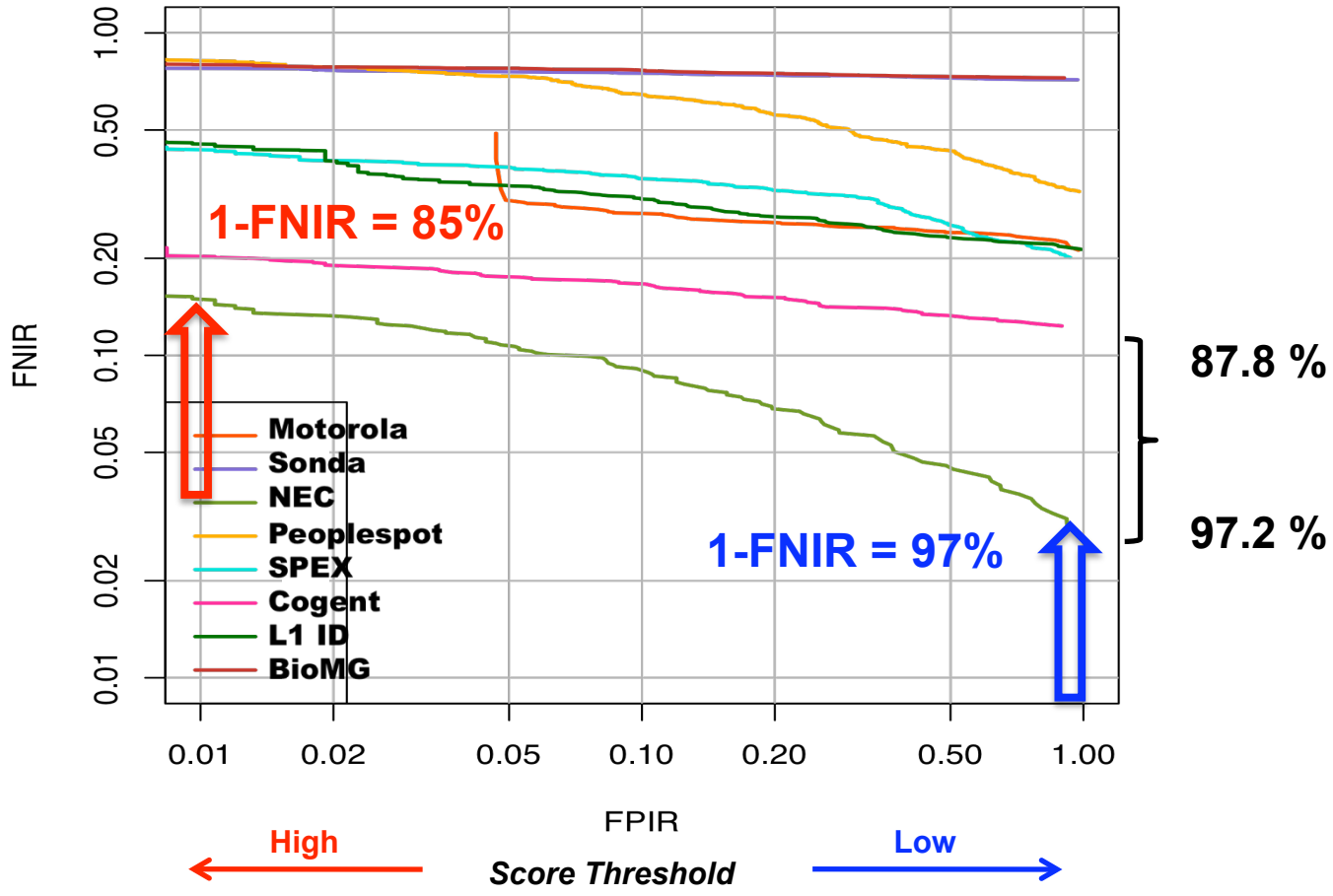
FPIR and FNIR

- FPIR = fraction of candidate lists which contain one or more non-mate entries after the original candidate list has been thresholded at score t and limited to length K.

- FNIR = fraction of candidate lists for which the enrolled mates do not appear in the top K positions with score greater than threshold, t. (aka “Miss Rate”)

DET – *match score*

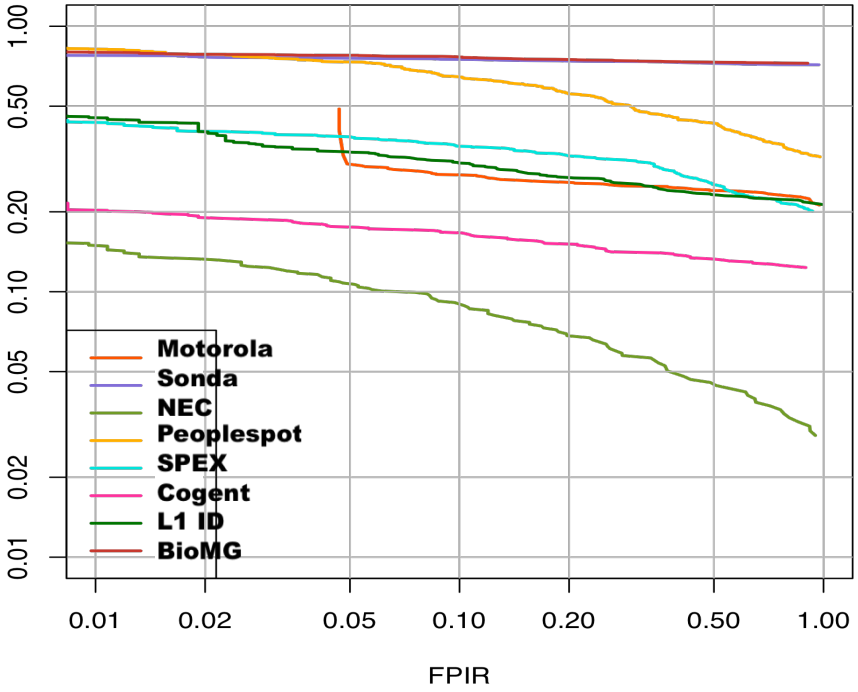
1000 ppi latents •100,000 fingerprint background



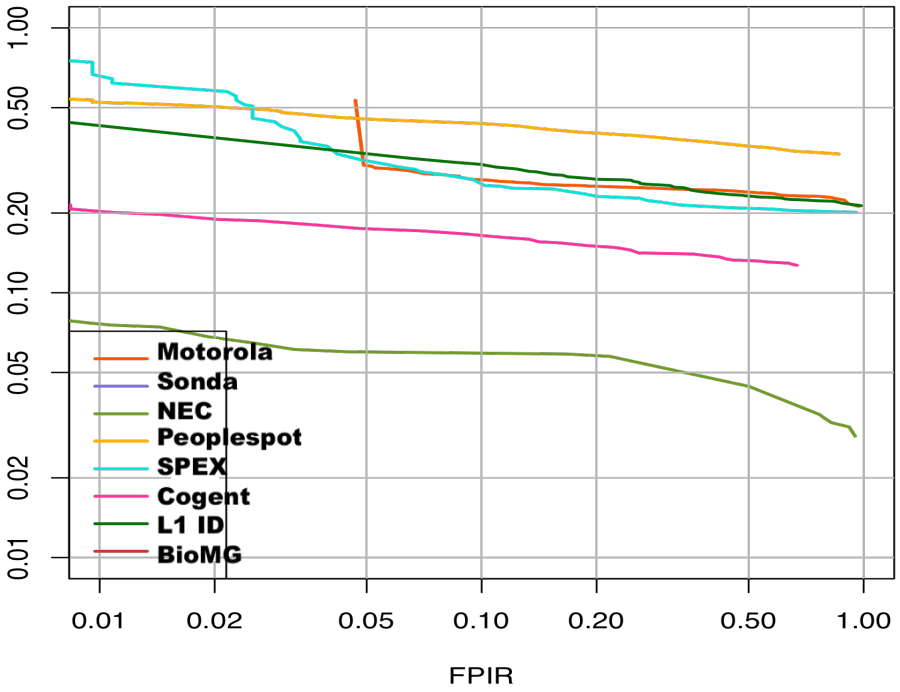
DET

1000 ppi latents •100,000 fingerprint background

match score

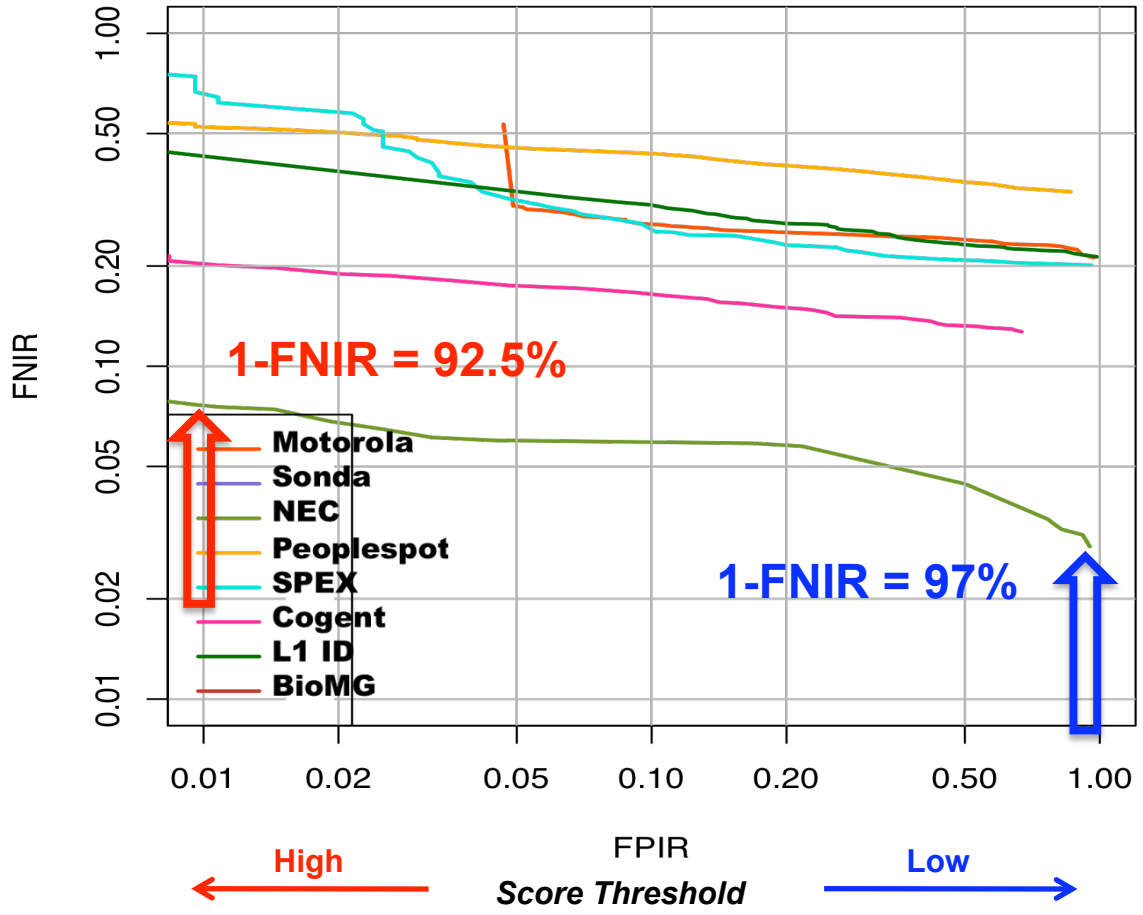


probability score



DET – *probability score*

1000 ppi latents •100,000 fingerprint background



Phase II Observations 1

- *Thresholding* based on an SDK provided *probability* score was shown to be more effective at reducing false-matches than the provided proprietary scores for two SDKs. This has important implications for candidate list reduction, interoperability, and fusion.
- A strong correlation exists between minutiae count and identification rate. Searches of latents with higher minutiae counts produced more accurate results.
- Candidate list fusion, using multi-fingers or multi-algorithms is a powerful mechanism for improving accuracy.

Phase II Observations 2

- The effect of increasing database size from 50,000 to 100,000 resulted in a one percentage-point average decrease in accuracy at rank 1
- The effect of resolution (1000 ppi vs. 500 ppi) was mixed and not statistically significant.
- The effect of region-of-interest itself was mixed, however, images with >50% area of ROI benefited the most.

Phase II Conclusions

1. Some matchers tested possess accuracies such that a limited class of latent fingerprints from operational casework can benefit from AFEM, thereby reducing some of the human workload during the AFIS latent fingerprint processes.
2. Specific measures (e.g. latent quality measures) do not currently exist for determining which latents are suitable for AFEM.
3. More testing is required to define AFEM limitations.

Caveats

- Latents and exemplars were identified by an operational AFIS
  - higher “quality” class of latents and exemplars
  - pre-selection and image processing for AFIS search
- Unconstrained processing time
- 500 ppi images not native scanning resolution (sub-sampled)
- AFEM accuracy is highly dependent on source, selection, and preparation of data. Study results may not be applicable to other datasets and operational databases.

For More Information…

Web ➔ http://fingerprint.nist.gov/latent

Email ➔ latent@nist.gov

Thank you!

Questions?